

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-030240
(43)Date of publication of application : 31.01.2003

(51)Int.Cl. G06F 17/30
G06F 12/00

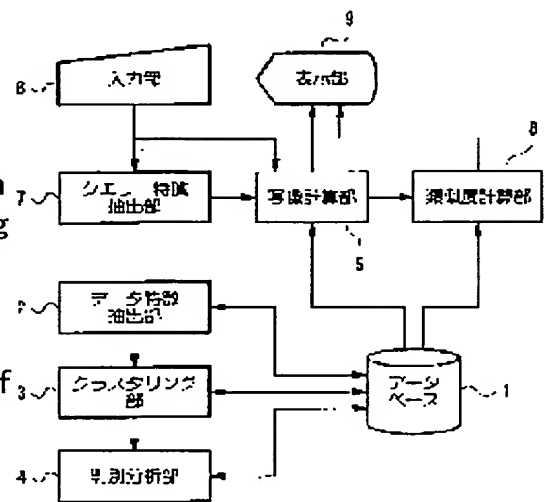
(21)Application number : 2001-214354 (71)Applicant : NTT DATA CORP
(22)Date of filing : 13.07.2001 (72)Inventor : SUENAGA TAKASHI
SAKANO EI

(54) DEVICE, METHOD AND PROGRAM FOR RETRIEVING DATA

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a data retrieving device capable of presenting data to a retriever in the form of making the correlation among data stored in a database easily understandable.

SOLUTION: A data retrieving device is provided with a means for extracting characteristic quantity having the number of dimensions being ≥ 4 dimensions from each piece of all data stored in the database, a means for dividing a plurality of pieces of data stored in the database into the prescribed number of clusters on the basis of characteristic quantity of the data, a means for calculating a projection queue for making the number of dimensions of the characteristic quantity of each piece of data ≤ 3 dimensions with respect to the data divided into the clusters by using a discriminant analysis, a means for extracting characteristic quantity having the number of dimensions being ≥ 4 dimensions from an inputted query, and a means for multiplying the data characteristic quantity and query characteristic quantity by the projection queue to calculate coordinates values the number of dimensions of which is ≤ 3 dimensions, and displaying the relation between the each piece of data stored in the database and the query with a scatter diagram by plotting the coordinates values.



LEGAL STATUS

[Date of request for examination] 13.07.2001
[Date of sending the examiner's decision of rejection] 19.10.2004
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's
decision of rejection]

[Date of requesting appeal against
examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] It is database retrieval equipment with which desired data are searched out of two or more data saved in the database. Said database retrieval equipment A data feature-extraction means to extract the characteristic quantity which has the number of dimension of 4-dimensional one or more from each of all the data saved in said database, The clustering means which divides into a predetermined number of clusters two or more data saved in said database based on the characteristic quantity of said data, A projection-matrix calculation means to compute the projection matrix for making the number of dimension of the characteristic quantity of each data below into a three dimension to the data injured by the cluster using discriminant analysis with said clustering means, A query feature-extraction means to extract the characteristic quantity which has the number of dimension of 4-dimensional one or more from the query inputted in order to search desired data, By calculating the coordinate value with which the multiplication of said projection matrix is carried out to the characteristic quantity of said data, and the characteristic quantity of said query, and a number of dimension becomes below a three dimension, and plotting this coordinate value Data retrieval equipment characterized by having a map count means to display the relation of the each data and the query which are saved in said database with a scatter diagram.

[Claim 2] Said data retrieval equipment is data retrieval equipment according to claim 1 characterized by to have further a similarity count means calculate and display each of data and the similarity of a query which were chosen by data selection means to choose the data located near the characteristic quantity of the query plotted by said map count means, and said data selection means based on the characteristic quantity of 4-dimensional one or more.

[Claim 3] Said similarity count means is data retrieval equipment according to claim 2 characterized by making into similarity Euclidean distance of the characteristic quantity which has the number of dimension of 4-dimensional one or more.

[Claim 4] It is the database search method which searches desired data out of two or more data saved in the database. Said database search method The data feature-extraction process in which the characteristic quantity which has the number of dimension of 4-dimensional one or more is extracted from each of all the data saved in said database, The clustering process which divides into a predetermined number of clusters two or more data saved in said database based on the characteristic quantity of said data, The projection-matrix calculation process which computes the projection matrix for making the number of dimension of the characteristic quantity of each data below into a three dimension to the data injured by the cluster using discriminant

analysis according to said clustering process, The query feature-extraction process in which the characteristic quantity which has the number of dimension of 4-dimensional one or more is extracted from the query inputted in order to search desired data, By calculating the coordinate value with which the multiplication of said projection matrix is carried out to the characteristic quantity of said data, and the characteristic quantity of said query, and a number of dimension becomes below a three dimension, and plotting this coordinate value The data retrieval approach characterized by having the map computation which displays the relation of the each data and the query which are saved in said database with a scatter diagram.

[Claim 5] Said data retrieval approach is the data retrieval approach according to claim 4 characterized by having further the similarity computation which calculates and displays each of data and the similarity of a query which were chosen by the data selection process which chooses the data located near the characteristic quantity of the query plotted by said map computation, and said data selection process based on the characteristic quantity of 4-dimensional one or more.

[Claim 6] Said similarity computation is the data retrieval approach according to claim 5 characterized by making into similarity Euclidean distance of the characteristic quantity which has the number of dimension of 4-dimensional one or more.

[Claim 7] It is the database retrieval program which searches desired data out of two or more data saved in the database. Said database retrieval program Data feature-extraction processing in which the characteristic quantity which has the number of dimension of 4-dimensional one or more is extracted from each of all the data saved in said database, The clustering processing which divides into a predetermined number of clusters two or more data saved in said database based on the characteristic quantity of said data, The projection-matrix calculation processing which computes the projection matrix for making the number of dimension of the characteristic quantity of each data below into a three dimension to the data injured by the cluster using discriminant analysis by said clustering processing, Query feature-extraction processing in which the characteristic quantity which has the number of dimension of 4-dimensional one or more is extracted from the query inputted in order to search desired data, By calculating the coordinate value with which the multiplication of said projection matrix is carried out to the characteristic quantity of said data, and the characteristic quantity of said query, and a number of dimension becomes below a three dimension, and plotting this coordinate value The data retrieval program characterized by making the map computation which displays the relation of the each data and the query which are saved in said database with a scatter diagram perform to a computer.

[Claim 8] Said data retrieval program is a data retrieval program according to claim 7 characterized by making the similarity computation which calculates and displays each of data and the similarity of a query which were chosen by the data selection processing which chooses the data located near the characteristic quantity of the query plotted by said map computation, and said data selection processing based on the characteristic quantity of 4-dimensional one or more perform to a computer further.

[Claim 9] Said similarity computation is a data retrieval program according to claim 8 characterized by making into similarity Euclidean distance of the characteristic quantity which has the number of dimension of 4-dimensional one or more.

* NOTICES *

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the data retrieval equipment and the data retrieval approach of showing the display of the retrieval result in a large scale database in the form which is easy to understand to a user, and a data retrieval program.

[0002]

[Description of the Prior Art] It is common that a data constellation most similar to the question (query) inputted in the data retrieval from a database is conventionally outputted in order of similarity. Since the data with which the data retrieval by this approach differed from the intention of a retrieval person are outputted in large quantities in many cases, a retrieval person has to input the question for narrowing down further, or has to discover desired data out of a lot of data, and has the problem that a retrieval person's burden is large and effectiveness is bad.

[0003] In order to solve such a problem, the approach of displaying the relation of data spatially is studied by displaying the relation of the data saved by carrying out a feature-vector expression which consists of two or more numeric values in the data saved in the database in the form of the scatter diagram of 2 which human being can grasp, or a three dimension. In this case, although the dimension reduction technique of searching for the coordinate of 2 or a three dimension is required in order to display on a scatter diagram when the number of dimension of the vector expressing data is four or more, principal component analysis has been conventionally used for this dimension reduction.

[0004]

[Problem(s) to be Solved by the Invention] However, in many databases, since there is much information required in order to express data, the feature vector expressing data turns into a vector-dimensional [thousands of], when many, at least 100 dimensions and. Although the number of data is about about ten pieces, namely, a scatter diagram effective for retrieval is generated by the display using the conventional principal component analysis in the small-scale database in which data are degenerating to the subspace of a low dimension, when the number of data exceeds hundreds, there is a problem that it is rare that a meaningful display is obtained and it is difficult to perform efficient retrieval which is the original purpose. Moreover, in having used principal component analysis, in order to perform the map to a low dimension for the purpose of saving all that distance relation about tens of thousands of [thousands which exist in high order former space to] feature vectors, all relation is spoiled little by little and distribution of the data mapped by the low dimension as a result serves as most or a

display which is not reflected at all in the distribution structure in a feature space. That is, since the display which saves the far and near relation in a feature space in low dimension space cannot be obtained when aimed at a large scale database as long as principal component analysis is used especially, there is a problem that retrieval effectiveness cannot be made high as a result.

[0005] This invention was made in view of such a situation, and aims at offering the data retrieval equipment which can be shown to a retrieval person in the form where the interrelation of the data saved in the database is intelligible, the data retrieval approach, and a data retrieval program.

[0006]

[Means for Solving the Problem] Invention according to claim 1 is database retrieval equipment with which desired data are searched out of two or more data saved in the database. Said database retrieval equipment A data feature-extraction means to extract the characteristic quantity which has the number of dimension of 4-dimensional one or more from each of all the data saved in said database, The clustering means which divides into a predetermined number of clusters two or more data saved in said database based on the characteristic quantity of said data, A projection-matrix calculation means to compute the projection matrix for making the number of dimension of the characteristic quantity of each data below into a three dimension to the data injured by the cluster using discriminant analysis with said clustering means, A query feature-extraction means to extract the characteristic quantity which has the number of dimension of 4-dimensional one or more from the query inputted in order to search desired data, By calculating the coordinate value with which the multiplication of said projection matrix is carried out to the characteristic quantity of said data, and the characteristic quantity of said query, and a number of dimension becomes below a three dimension, and plotting this coordinate value It is characterized by having a map count means to display the relation of the each data and the query which are saved in said database with a scatter diagram.

[0007] Invention according to claim 2 is characterized by to equip said data retrieval equipment with a similarity count means to calculate and display each of data and the similarity of a query which were chosen by data selection means to choose the data located near the characteristic quantity of the query plotted by said map count means, and said data selection means based on the characteristic quantity of 4-dimensional one or more, further.

[0008] It is characterized by invention according to claim 3 making similarity Euclidean distance of the characteristic quantity in which said similarity count means has the number of dimension of 4-dimensional one or more.

[0009] Invention according to claim 4 is a database search method which searches desired data out of two or more data saved in the database. Said database search method The data feature-extraction process in which the characteristic quantity which has the number of dimension of 4-dimensional one or more is extracted from each of all the data saved in said database, The clustering process which divides into a predetermined number of clusters two or more data saved in said database based on the characteristic quantity of said data, The projection-matrix calculation process which computes the projection matrix for making the number of dimension of the characteristic quantity of each data below into a three dimension to the data injured by the cluster using discriminant analysis according to said clustering process, The query feature-extraction process in which the characteristic quantity which has the number of dimension of 4-dimensional one or more is extracted from the query

inputted in order to search desired data, By calculating the coordinate value with which the multiplication of said projection matrix is carried out to the characteristic quantity of said data, and the characteristic quantity of said query, and a number of dimension becomes below a three dimension, and plotting this coordinate value It is characterized by having the map computation which displays the relation of the each data and the query which are saved in said database with a scatter diagram.

[0010] It is characterized by having further the similarity computation which invention according to claim 5 calculates each of data and the similarity of a query which were chosen by the data selection process which chooses the data located near [where said data retrieval approach was plotted by said map computation] the characteristic quantity of a query, and said data selection process based on the characteristic quantity of 4-dimensional one or more, and is displayed.

[0011] Invention according to claim 6 is characterized by said similarity computation making similarity Euclidean distance of the characteristic quantity which has the number of dimension of 4-dimensional one or more.

[0012] Invention according to claim 7 is a database retrieval program which searches desired data out of two or more data saved in the database. Data feature-extraction processing in which the characteristic quantity which has the number of dimension of 4-dimensional one or more is extracted from each of all the data with which said database retrieval program is saved in said database, The clustering processing which divides into a predetermined number of clusters two or more data saved in said database based on the characteristic quantity of said data, The projection-matrix calculation processing which computes the projection matrix for making the number of dimension of the characteristic quantity of each data below into a three dimension to the data injured by the cluster using discriminant analysis by said clustering processing, Query feature-extraction processing in which the characteristic quantity which has the number of dimension of 4-dimensional one or more is extracted from the query inputted in order to search desired data, By calculating the coordinate value with which the multiplication of said projection matrix is carried out to the characteristic quantity of said data, and the characteristic quantity of said query, and a number of dimension becomes below a three dimension, and plotting this coordinate value It is characterized by making the map computation which displays the relation of the each data and the query which are saved in said database with a scatter diagram perform to a computer.

[0013] Invention according to claim 8 is characterized by said data retrieval program making the similarity computation which calculates and displays each of data and the similarity of a query which were chosen by the data selection processing which chooses the data located near the characteristic quantity of the query plotted by said map computation, and said data selection processing based on the characteristic quantity of 4-dimensional one or more perform to a computer further.

[0014] Invention according to claim 9 is characterized by said similarity computation making similarity Euclidean distance of the characteristic quantity which has the number of dimension of 4-dimensional one or more.

[0015]

[Embodiment of the Invention] Hereafter, the data retrieval equipment by 1 operation gestalt of this invention is explained with reference to a drawing. Drawing 1 is the block diagram showing the configuration of this operation gestalt. In this drawing, a sign 1 shall be a large scale database for retrieval, and document data shall be saved here. A sign 2 is the data feature-extraction section which extracts the description of data by

changing into high order dimension numeric-value vector data each data saved in the database 1. A sign 3 is the clustering section which clusters to the high order dimension numeric-value vector data saved in the database 1. A sign 4 is the discriminant analysis section which performs discriminant analysis to the clustered high order dimension numeric-value vector data. A sign 5 is the map count section which maps the high order former vector which expresses data using the map obtained by discriminant analysis into a low dimension. A sign 6 is the query input section which inputs a question (henceforth a query), and consists of keyboards etc. In addition, the input section 6 may be the configuration of reading a data file. A sign 7 is the query feature-extraction section which extracts the description of the query inputted from the input section 6. A sign 8 is the similarity count section which calculates the similarity of a query and data. A sign 9 is a display which consists of CRT, a liquid crystal display, etc.

[0016] Here, the principle of the data retrieval of this invention is briefly explained with reference to drawing 5 and 6. The thing for carrying out the two-dimensional expression of the numerical vector of high order origin to do for dimension reduction is the purpose so that human being may tend to recognize intuitively the data aggregate to which this invention is similar. Here, since it is easy, it makes to carry out the two-dimensional expression of the numerical vector of a three dimension into an example, and explains. (a) of drawing 5 is drawing which expressed the feature vector of each data at the point. In this drawing, the point that distance is near considers that it is similar data, and carries out the cluster division of each data using k method of averaging. And if the two-dimensional flat surface shown with the broken line of drawing 5 (a) is searched for by discriminant analysis and each point is mapped at this two-dimensional flat surface, drawing as shown in (c) of drawing 5 will be obtained. If the query which is equivalent to retrieval conditions to this two-dimensional flat surface is plotted, the assembly of the data near conditions can be known intuitively.

[0017] It is difficult to, recognize the assembly of the data which are similar as shown in (b) of drawing 5, when the two-dimensional flat surface which takes a map is not suitable on the other hand. It is the purpose that this invention searches for efficiently the two-dimensional flat surface which the characteristic quantity of a dimension is not lost as much as possible when data are the numerical vector of high order origin, and can recognize the assembly of similar data intuitively. For this reason, this invention searches for the two-dimensional flat surface where the mapping point on the two-dimensional flat surface based on clusters and the variance of each data become equal using the method of discriminant analysis, after carrying out the cluster division of each data, as shown in drawing 6. By doing in this way, the data which bring together the data belonging to a certain cluster in near, and belong to a different cluster are separated, and it becomes possible to display to a data retrieval person.

[0018] Next, retrieval actuation of the data retrieval equipment shown in drawing 1 is explained. Here, from the document saved in the database 1, actuation which searches a similar document is made into an example, and it explains in order to search whether there was any inquiry similar in the past to the inquiry electronic mail in a help desk. Off-line processing before performing introduction and data retrieval is explained. First, the data feature-extraction section 2 reads the document data saved in the database 1 (step S1). And it asks for the feature vector x_n ($n=1 \dots N$) of the read document data (step S2). This feature vector x_n is called for based on the histogram of each frequency of occurrence of two or more words required for data retrieval, in a database 1, is related with document data and saved.

[0019] A word required for the data retrieval for which it opts beforehand For example, a "computer", It is defined as "cooperation", the "display", and the "keyboard". The target document data When saying, "A computer operates by cooperating with the equipment of not only the computer itself but a perimeter", since the number of one piece, a "display", and "keyboards" is zero, respectively for two pieces and "cooperation", a "computer" This document is expressed as the combination of two or more numeric values called (2, 1, 0, 0), i.e., a vector, is related with document data and saved. This processing is performed to all the document data saved in the database 1, and will be in the condition that the feature vector was associated and saved for every document data in the database 1 at this time. Then, the data feature-extraction section 2 notifies that the feature extraction was completed to the clustering section 3.

[0020] Next, the clustering section 3 sets k (k is the two or more natural numbers) individual ejection and this k document data for the document data saved in the database 1 as a temporary cluster core at random (step S3), and gives each the cluster number of $1 - k$. Then, the clustering section 3 reads in order the document data saved in the database 1. And out of k document data taken out previously, the read document data ask for the nearest document data, and give the cluster number of the nearest document data temporarily to the read document data. It is a thing here with the nearest Euclidean distance of a feature vector in it being the nearest. This processing is performed to all document data. It means that the cluster number of either $1 - k$ is temporarily given by this, and document data were classified into k clusters according to it to all document data.

[0021] Next, the clustering section 3 calculates the average of the subset of the document data belonging to each cluster, and sets this average as a new cluster core (step S5). And the clustering section 3 repeats step S4 and processing of S5 until a new cluster core becomes the same as the last cluster core (step S6), it adds them to document data by using as a label the cluster number temporarily given to each document data, and is saved in a database 1 (step S7). Then, the clustering section 3 notifies that clustering processing was completed to the discriminant analysis section 4.

[0022] Next, the discriminant analysis section 4 calculates the total average m of the feature vector of the document data of N individual saved in the database 1 (step S8). Then, the discriminant analysis section 4 calculates the average m_i of each cluster $1 - k$ (step S9). And the discriminant analysis section 4 calculates the cluster internal variance matrix SW and the variance matrix SB between clusters (steps S10 and S11). And the discriminant analysis section 4 solves the eigenvalue problem of $SW-1SB$ (step S12). That is, the solution to which the distance of each cluster becomes far, and the distance of each data in a cluster becomes near is calculated.

[0023] Next, the discriminant analysis section 4 rearranges into descending the characteristic value acquired in step S12 (step S13), and takes out characteristic vector W corresponding to the 1st and the 2nd characteristic value (step S14). And by carrying out matrix operation to all the document data saved in the database 1, Coordinate y_n is calculated (step S15) and the result is saved in a database 1. Thus, a scatter diagram will be obtained, if it means that the document data saved in the database 1 are divided into k clusters, and it was changed into the coordinate y_n whose vector data of high order origin is data in which a two-dimensional expression is possible and this coordinate y_n is plotted by off-line processing of steps S1-S15. In addition, off-line processing of steps S1-S15 shown in drawing 2 is periodically

performed to compensate for document data newly being saved in a database 1.

[0024] Next, the actuation which searches desired data from the document data with which off-line processing mentioned above was performed is explained. First, an operator will input this mail as a query, if e-mail reaches a help desk (step S21). The input section 6 reads the contents of this mail, and outputs those contents to the query feature-extraction section 7. In response, the query feature-extraction section 7 divides the contents of e-mail into a word, evaluates them with the frequency of occurrence of that word, asks for feature-vector u (step S22), and outputs this feature-vector u to the map count section 5.

[0025] Next, the map count section 5 calculates the coordinate v of a query using the projection matrix (characteristic vector) W for which it asked in step S14 mentioned above (step S23). And the map count section 5 displays the coordinate v searched for on a display 9. Moreover, the map count section 5 reads the data (coordinate y_n searched for in step S15) of the scatter diagram saved in the database 1, and piles up and displays them on the screen which displayed the coordinate v of a query (step S24). This screen is seen and an operator chooses by the input section 6 by making the data near a query applicable to retrieval. An example of the scatter diagram displayed on a display 9 at this time is shown in drawing 4. The document data with which the word by which drawing 4 was defined beforehand was saved in 2000 words and a database are an example of processing activation in case the value of 500 pieces and k is 6. In drawing 4, each data with which the sunspot was plotted based on Coordinate y_n is expressed, and Sign Q expresses the query plotted based on Coordinate v . Moreover, Sign A shows the area which the operator chose.

[0026] The map count section 5 considers that the data which exist in this area are data similar to a query, and notifies the data which exist in this area to the similarity count section 8. In response, the similarity count section 8 calculates similarity by making applicable to retrieval only the data notified from the map count section 5, and displays the result on a display 9. With similarity here, it is considered that similarity is high in order with the near Euclidean distance of the feature vector of the high order origin for which it asked in step S2. If this similarity chooses high document data, the document data near the contents of mail of a query are discoverable.

[0027] Thus, since the scatter diagram expressed the relation of the data and the query which are saved in the database 1, it becomes possible to raise the effectiveness of retrieval of a retrieval person by observing the data located near the query. This can be applied also to design assistance of a pattern recognition dictionary, and character recognition and speech recognition, and the application to the data-mining technique used by customer relationship management (Customer Relationship Management) etc. is also still more possible for it.

[0028] in addition -- although k method of averaging was made into the example and explained as an approach of clustering in the explanation mentioned above -- Word -- it is also possible to use the clustering technique, such as law. Moreover, although it was made to carry out the two-dimensional expression of the relation of data, you may make it a three-dimension expression express the relation of data in drawing 4. What is necessary is just to take out characteristic vector $W = (w_1 w_2 w_3)$ to the 1-3rd characteristic value in step S14 at this time.

[0029] In addition, the program which recorded on the record medium which can computer read the program for realizing the function of each processing shown in drawing 2 and 3, and was recorded on this record medium may be made to read into a computer system, and data retrieval processing may be performed by performing. In

addition, hardware, such as OS and a peripheral device, shall be included with a "computer system" here. Moreover, if a "computer system" is the case where the WWW system is used, it shall also include a homepage offer environment (or display environment). Moreover, "the record medium in which computer reading is possible" means storage, such as a hard disk built in portable media, such as a flexible disk, a magneto-optic disk, ROM, and CD-ROM, and a computer system. Furthermore, the thing holding a fixed time amount program shall also be included ["whose record medium in which computer reading is possible" is] like the volatile memory (RAM) inside the computer system used as a server when a program is transmitted through communication lines, such as networks, such as the Internet, and the telephone line, or a client.

[0030] Moreover, the above-mentioned program may be transmitted to other computer systems through a transmission medium from the computer system which stored this program in storage etc. by the carrier wave in a transmission medium. Here, the "transmission medium" which transmits a program says the thing of a medium which has the function to transmit information like communication lines (communication wire), such as networks (communication network), such as the Internet, and the telephone line. Moreover, the above-mentioned program may be for realizing a part of function mentioned above. Furthermore, you may be what can realize the function mentioned above in combination with the program already recorded on the computer system, and the so-called patch file (difference program).

[0031]

[Effect of the Invention] Since it becomes possible to map to the low dimension space which can be understood to human being, with the spatial relations of the data expressed as a high order former vector saved according to this invention as explained above, the effectiveness that the retrieval effectiveness of a database improves is acquired.

[Translation done.]

* NOTICES *

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the block diagram showing the configuration of 1 operation gestalt of this invention.

[Drawing 2] It is the flow chart which shows actuation of the data retrieval equipment shown in drawing 1 .

[Drawing 3] It is the flow chart which shows actuation of the data retrieval equipment shown in drawing 1 .

[Drawing 4] It is the explanatory view showing an example of the screen displayed on a display 9.

[Drawing 5] It is an explanatory view for explaining the data retrieval principle by this invention.

[Drawing 6] It is an explanatory view for explaining the data retrieval principle by this invention.

[Description of Notations]

1 ... Database 2 [4 / 6 / ... The similarity count section, 9 / ... Display / ... The input section, 7 ... The query feature-extraction section, 8 / ... The discriminant analysis section, 5 ... Map count section] ... The data feature-extraction section, 3 ... Clustering section

[Translation done.]

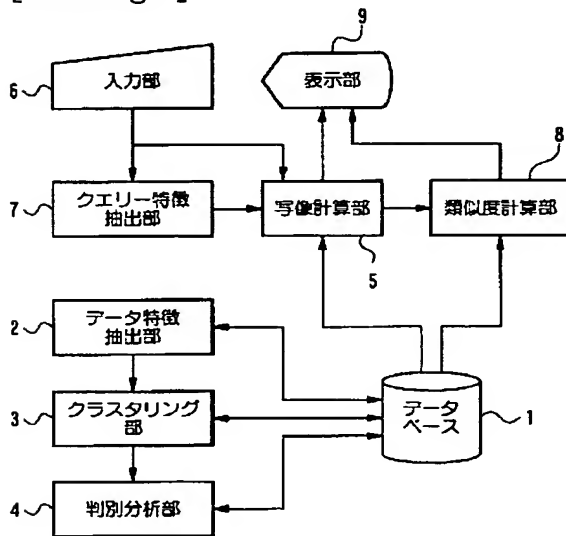
* NOTICES *

JPO and NCIP are not responsible for any damages caused by the use of this translation.

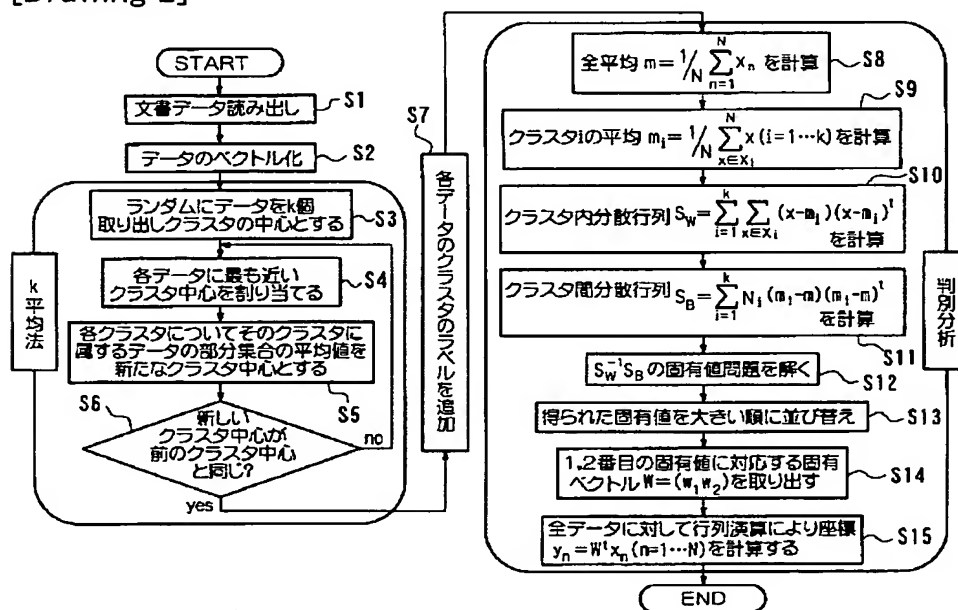
- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DRAWINGS

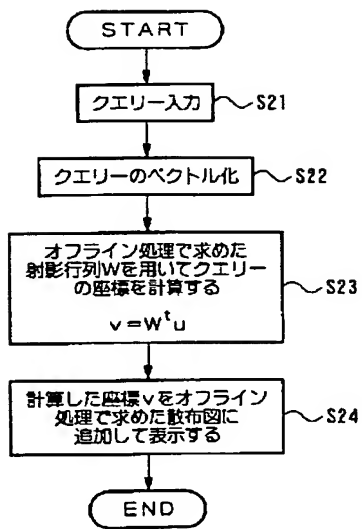
[Drawing 1]



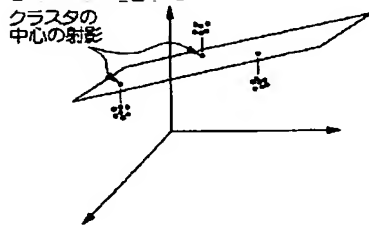
[Drawing 2]



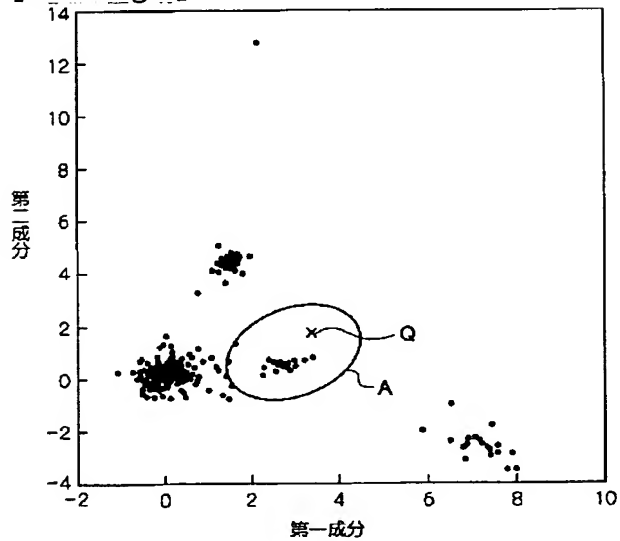
[Drawing 3]



[Drawing 6]



[Drawing 4]



[Drawing 5]

